Chapter 15
PIM Overview

Protocol Independent Multicast (PIM) is used for efficiently routing to multicast groups that might span wide-area and interdomain internetworks. It is called “protocol independent” because it does not depend on a particular unicast routing protocol. The JUNOS software supports sparse mode, dense mode, and sparse-dense mode.

PIM Standards

The JUNOS PIM implementation complies with the following documents:

RFC 2362, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*

Protocol Independent Multicast Version 2 Dense Mode Specification, Internet draft draft-ietf-pim-v2-dm-03.txt

Anycast RP Mechanism Using PIM and MSDP, Internet draft draft-ietf-mboned-anycast-rp-05.txt

Bootstrap Router (BSR) Mechanism for PIM Sparse Mode, Internet draft draft-ietf-pim-sm-bsr-02.txt

RFC 2547, *BGP/MPLS VPNs*

Multicast in MPLS/BGP VPNs, Section 2 (Multicast Domains) of Internet draft draft-rosen-vpn-mcast-00.txt

To access Internet RFCs and drafts, go to the IETF Web site at http://www.ietf.org.

PIM Modes

Because the mode you choose determines the PIM configuration properties, you first must decide whether PIM will operate in sparse, dense, or sparse-dense mode in your network. Each mode has distinct operating advantages in different network environments.

In sparse mode, routers must join and leave multicast groups explicitly. Upstream routers do not forward multicast traffic to a router unless it has sent an explicit request (by means of a Join message) to the RP router to receive this traffic. The RP serves as the root of the shared multicast delivery tree and is responsible for forwarding multicast data from different sources to the receivers.

Sparse mode is well suited to the Internet, where frequent interdomain Joins and Prunes are common.

Unlike sparse mode, in which data is forwarded only to routers sending an explicit PIM Join request, dense mode implements a *flood-and prune* mechanism, similar to the Distance Vector Multicast Routing Protocol (DVMRP). In dense mode, a router receives the multicast data on the incoming interface (IIF), then forwards the traffic to the outgoing interface (OIL). Flooding occurs periodically, and is used to refresh state information, such as the source IP address and multicast group pair. If the router has no interested receivers for the data, and the OIL becomes empty, the router sends a PIM Prune message upstream.

Dense mode works best in networks where few or no Prunes occur. In such instances, dense mode is actually more efficient than sparse mode.

Sparse-dense mode, as the name implies, allows the interface to operate on a per-group basis in either sparse or dense mode. A group specified as “dense” is not mapped to an RP. Instead, data packets destined for that group are forwarded by means of PIM-DM rules. A group specified as “sparse” is mapped to an RP, and data packets are forwarded by means of PIM-SM rules.

Sparse-dense mode is useful in networks implementing auto-RP for PIM-SM.

For more information about how the PIM modes operate, see:

PIM Sparse Mode on page 93

PIM Dense Mode on page 108

PIM Sparse-Dense Mode on page 109

For more information about mode-dependent configurations, see:

Configure PIM Dense Mode Properties on page 118

Configure PIM Sparse Mode Properties on page 118

Configure Sparse-Dense Mode Properties on page 126

PIM Sparse Mode

A PIM sparse mode (PIM-SM) domain uses Reverse Path Forwarding (RPF) to create a path from a data source to the receiver requesting the data. When a receiver issues an explicit Join request, an RPF check is triggered. A (*,G) PIM Join message is sent towards the RP from the receiver's DR. (By definition, this message is actually called a Join/Prune message, but for clarity in this description, it is called either Join or Prune, depending on its context.) The Join message is multicast hop by hop upstream to the ALL-PIM-ROUTERS group (224.0.0.13) by means of each router's RPF interface until it reaches the RP. The RP router receives the (*,G) PIM Join message and adds the interface on which it was received to the OIL of the RPT forwarding state entry. This builds the rendezvous point tree (RPT) connecting the receiver with the RP. The RPT remains standing, even if no active sources generate traffic.

Note

State—the (*,G) or (S,G) entries—is the information used for forwarding unicast or multicast packets. S is the source IP address, G is the multicast group address, and * represents any source sending to group G. Routers keep track of the multicast forwarding state for the incoming and outgoing interfaces for each group.

When a source becomes active, the source's DR encapsulates multicast data packets into a PIM Register message and sends them by means of unicast to the RP router.

If the RP router has interested receivers in the PIM-SM domain, the RP router sends a PIM Join message towards the source to build a shortest path tree (SPT) back to the source. The source sends multicast packets out on the LAN, and the source's DR encapsulates the packets in a PIM Register message and forwards it towards the RP router by means of unicast. The RP router receives PIM Register messages back from the source, and thus adds a new source to the distribution tree, keeping track of sources in a PIM table. Once an RP router receives packets natively (with S,G), it sends a Register Stop to stop receiving the register messages by means of unicast.

In actual application, many receivers with multiple SPTs are involved in a multicast traffic flow. To simply illustrate the process, we'll follow the multicast traffic from the RP router to one receiver. In such a case, the RP router begins sending multicast packets down the RPT towards the receiver's DR for delivery to the interested receivers. When the receiver's DR gets the first packet from the RPT, the DR sends a PIM Join message towards the source's DR to start building an SPT back to the source. When the source's DR receives the PIM Join from the receiver's DR, it starts sending traffic down all SPTs. When the first multicast packet is received by the receiver's DR, the receiver's DR sends a PIM Prune message to the RP router to stop duplicate packets being sent through the RPT. In turn, the RP router stops sending multicast packets to the receiver's DR, and sends a PIM Prune message for this source over the RPT toward's the source's DR to halt multicast packet delivery to the RP router from that particular source.

If the RP router receives a PIM register message from an active source, but has no interested receivers in the PIM-SM domain, it still adds the active source into the PIM table. However, after adding the active source into the PIM table, the RP router sends a Register Stop message. The RP router knows of the active source's existence and no longer needs to receive advertisement of the source (which utilizes resources).

This section contains more information about the routers and PIM-SM functions briefly described above:

Designated Router

When discussing a PIM-SM domain, there are two types of designated routers to consider:

- The receiver's designated router (DR) sends PIM Join and PIM Prune messages from the receiver network toward the RP.
- The source's designated router (DR) sends Register messages from the source network to the RP.

Regardless of whether it is the receiver's DR or the source's DR, a DR is selected from other routers in a network by the exchange of IP addresses. Neighboring PIM-SM routers multicast periodic Hello messages to each other every 30 seconds (the default). Upon receipt of a Hello message, a router stores the IP address and priority for that neighbor. If the DR priorities match, the router with the highest IP address is selected as the DR.

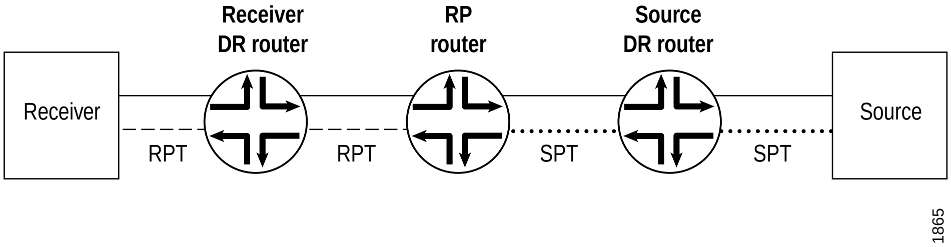
If a DR fails, a new one is selected using the same process of comparing IP addresses.

Rendezvous Point

The rendezvous point (RP) router serves as the information exchange point for the other routers. All routers in a PIM domain must provide mapping to an RP router. It is the only router that needs to know the active sources for a domain—the other routers just need to know how to get to the RP. In this way, the RP matches receivers with sources.

The RP router is downstream from the source and forms one end of the SPT. The RP router is upstream from the receiver, and thus, forms one end of the RPT (see Figure 11).

Figure 11: The RP as Part of the RPT and SPT



The benefit of using the RP as the information exchange point is that it reduces the amount of state in non-RP routers. No network flooding is required to provide non-RP routers information about active sources.

RP Mapping Options

RPs can be learned by one of the following mechanisms:

Static configuration (recommended)

Auto-RP

Bootstrap router

Static Configuration

You can configure a static RP configuration that is very similar to static routes. A static configuration has the benefit of operating in PIM version 1 or version 2. When configuring the static RP, the RP address you select for a particular group must be consistent across all routers in a multicast domain.

A static configuration is simple and convenient. However, if the statically defined RP router become unreachable there is no automatic failover to another RP router. To remedy this problem, you can use anycast RP. The anycast RP configuration is multiple RPs with a static configuration using the same IP address and also using MSDP. When MSDP is used along with PIM-SM, anycast RP provides a faster failover rate than auto-RP or a bootstrap router.

For configuration information, see “Configure Static RPs” on page 122. For anycast RP information, see “RP Mapping Options” on page 95.

Auto-RP

You can configure a more dynamic way of assigning RPs in a multicast network by means of auto-RP. When you configure auto-RP for a router, the router learns the address of the RP in the network automatically and has the added advantage of operating in PIM version 1 and version 2.

Although auto-RP is a non-standard (non-RPF-based) function that typically uses dense mode PIM to advertise control traffic, it provides an important failover advantage that simple static RP assignment does not. You can configure multiple routers as RP candidates. Should the elected RP stop operating, one of the other preconfigured routers takes over the RP functions. This capability is controlled by the auto-RP mapping agent.

For more information, see “Configure Auto-RP Announcement and Discovery” on page 123.

Bootstrap Router

To determine which router is the RP, all routers within a PIM-SM domain collect bootstrap messages. A PIM-SM domain is a group of routers that all share the same RP router. The domain’s bootstrap router originates bootstrap messages, which are sent hop by hop within the domain. The routers use bootstrap messages to distribute RP information dynamically and to elect a bootstrap router when necessary.

For more information, see “Configure Bootstrap Properties” on page 122.

Building an RPT between RP and Receivers

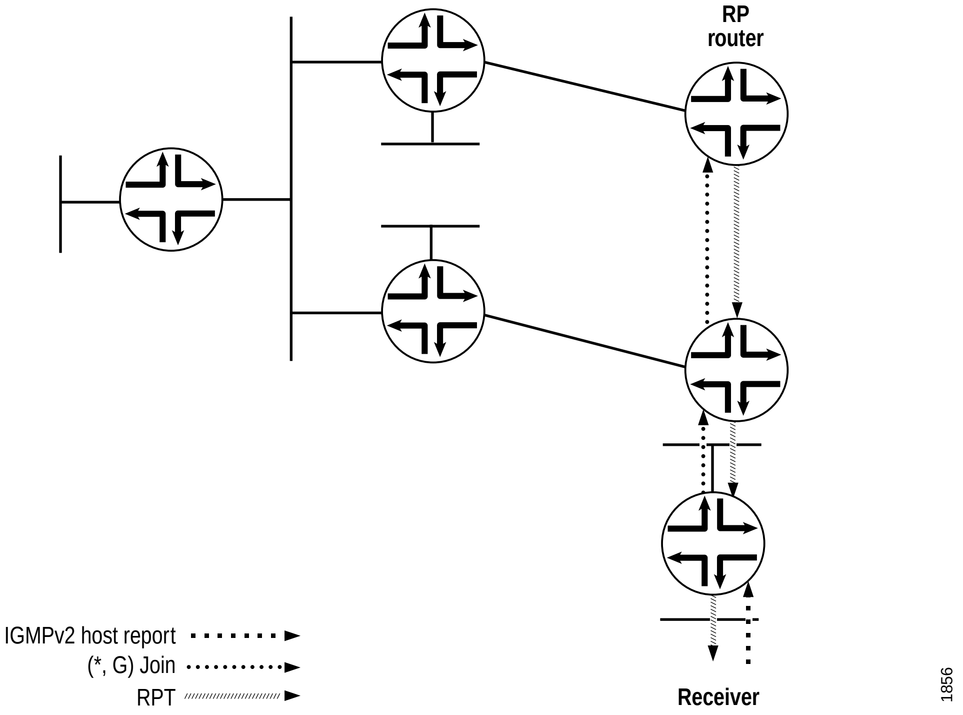
The RPT is the path between the RP and receivers (hosts) in a multicast group (see Figure 12). The RPT is built using a PIM Join message from a receiver’s DR:

1. A receiver announces a desire to join group (G) with an IGMP Host Membership Report. A PIM-SM router, the receiver’s DR, receives the report on a directly attached subnet and creates an RPT branch for the multicast group of interest.

2. The receiver’s DR sends a Pim Join message to its RPF neighbor, the next-hop address in the RPF table or the unicast routing table.

3. The PIM Join message travels up the tree, multicast to the ALL-PIM-ROUTERS group (224.0.0.13). Each router in the tree finds its RPF neighbor using either the RPF table or the unicast routing table. This is done until the message reaches the RP and forms the RPT. Routers along the path set up multicast forwarding state to forward requested multicast traffic back down the RPT to the receiver.

Figure 12: Building an RPT between RP and Receiver



PIM-SM Source Registration

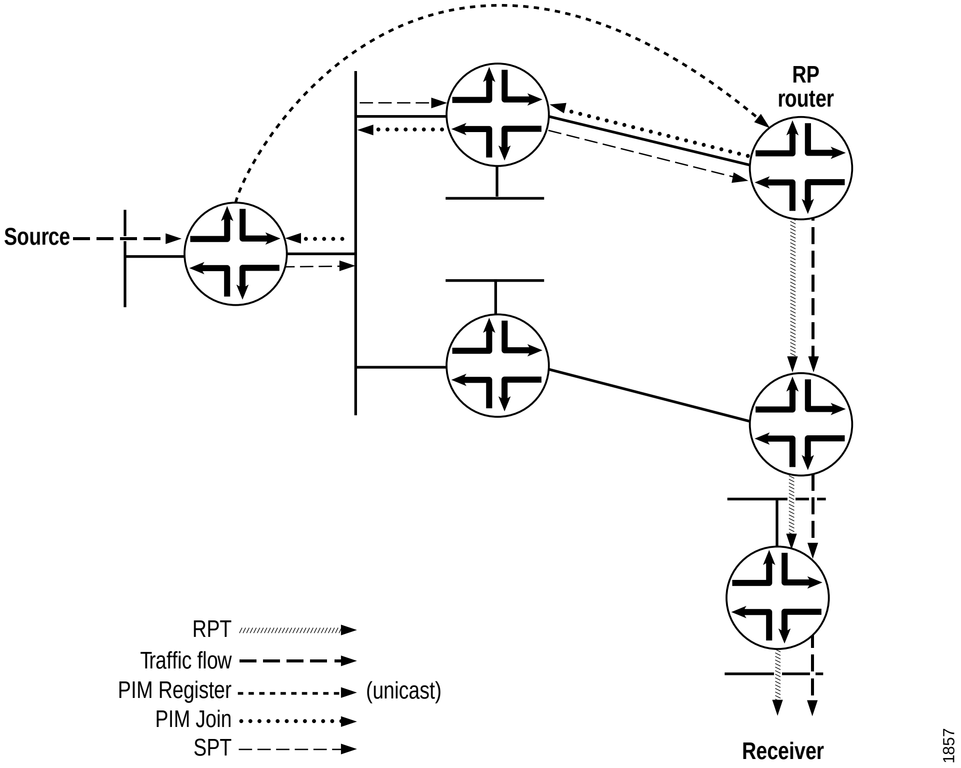
The RPT is a unidirectional tree, permitting traffic to flow down from the RP to the receiver in one direction. For multicast traffic to reach the receiver from the source, another branch of the distribution tree called the SPT needs to be built from the source's DR to the RP.

The SPT is created in the following way:

1. The source becomes active, sending out multicast packets on the LAN to which it is attached. The source's DR receives the packets and encapsulates them in a PIM Register message, which it sends out to the RP router. (see Figure 13).

2. When the RP router receives the PIM Register message from the source, it sends a PIM Join message back to the source.

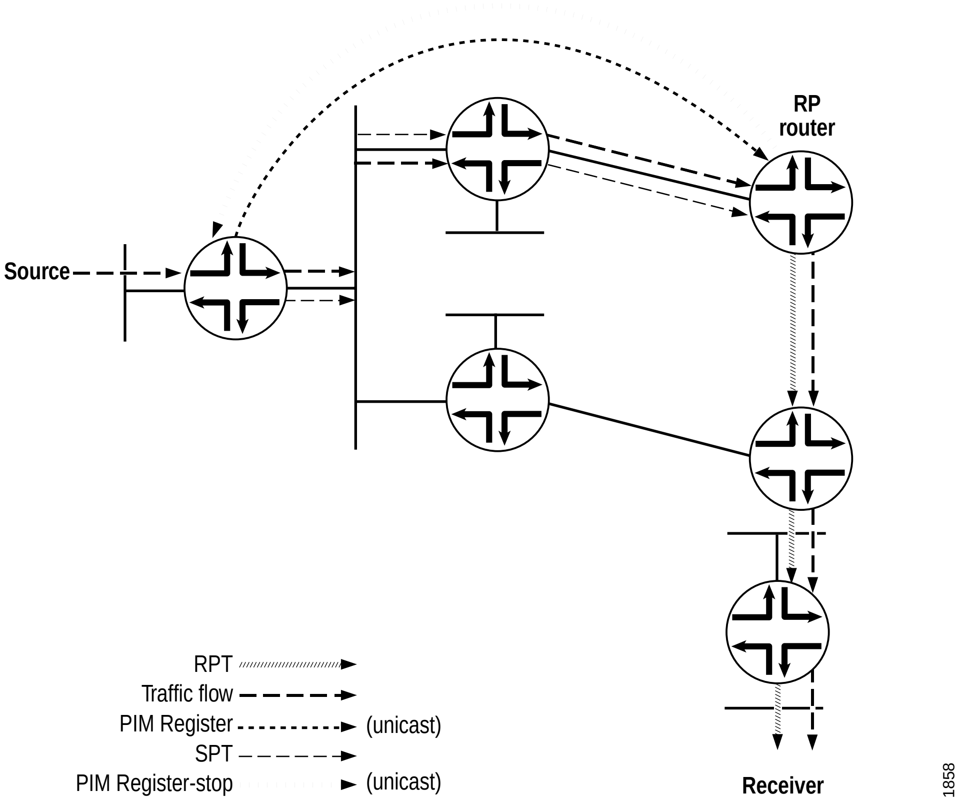
Figure 13: PIM Register Message and PIM Join Message Exchanged



3. The source’s DR receives the PIM Join message, and begins sending traffic down the SPT towards the RP router (see Figure 14).

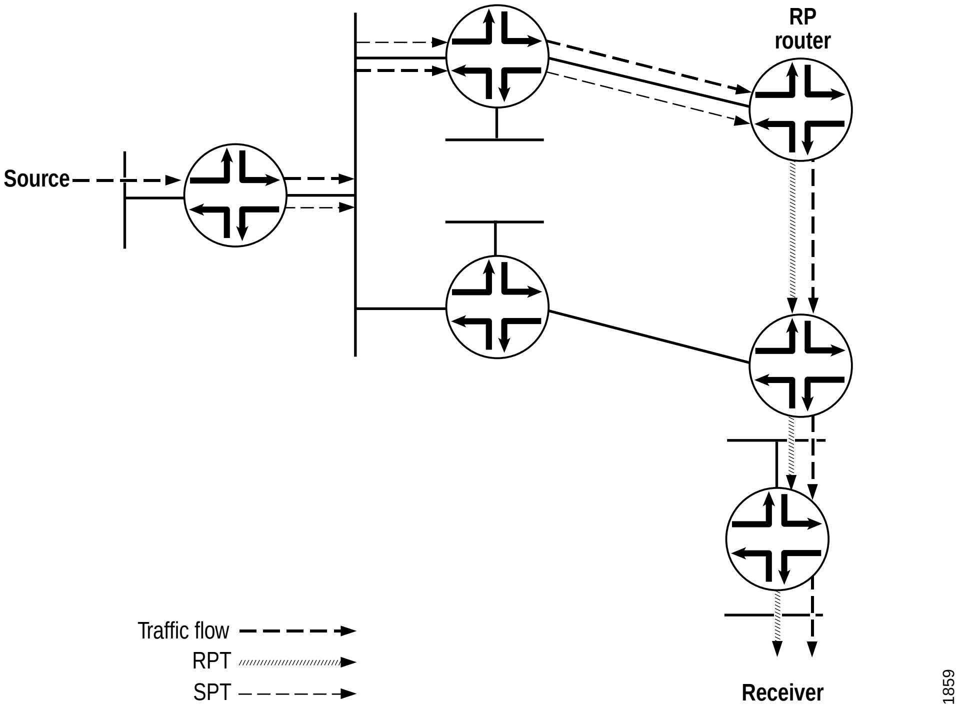
4. Once traffic is received by the RP router, it sends a Register-Stop message to the source’s DR to stop the register process.

Figure 14: Traffic Sent from the Source to the RP Router



5. The RP router sends the multicast traffic down the RPT towards the receiver (see Figure 15).

Figure 15: Traffic Sent from the RP Router Towards the Receiver



PIM-SM Shortest Path Tree (SPT) Cutover

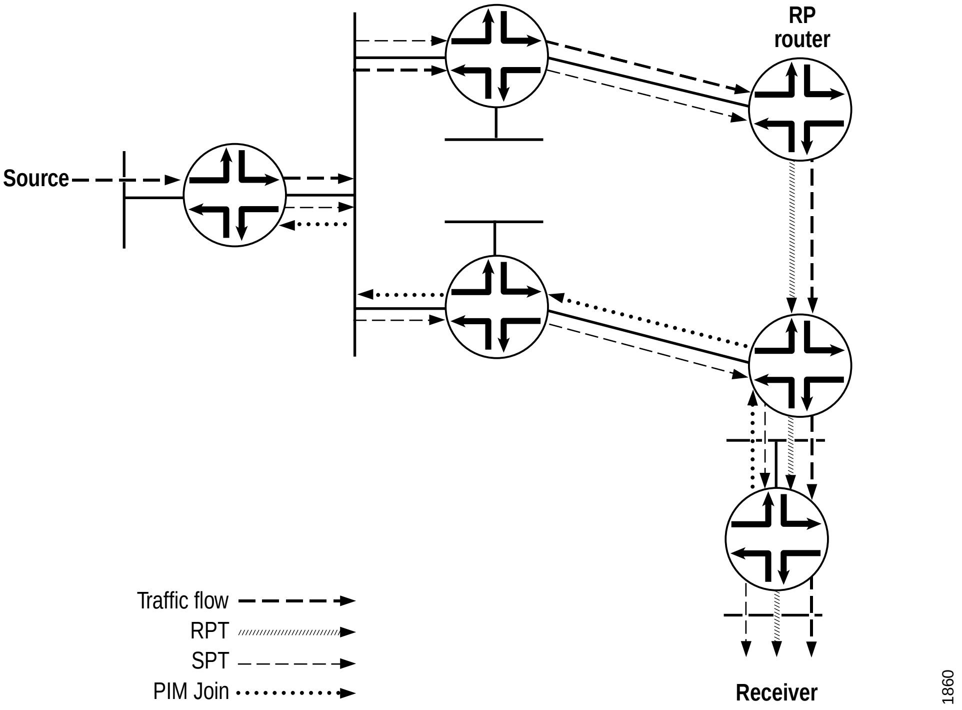
The problem of using the RPT to deliver multicast traffic to the receiver is that it may not be the most direct path. Instead of continuing to use the SPT to the RP and the RPT towards the receiver, a direct SPT is created between the source and the receiver in the following way:

1. Once the receiver's DR receives the first multicast packet from the source, the DR sends a PIM Join message to its RPF neighbor (see Figure 16).

2. The source's DR receives the PIM Join message, and an additional (S,G) state is created to form the SPT.

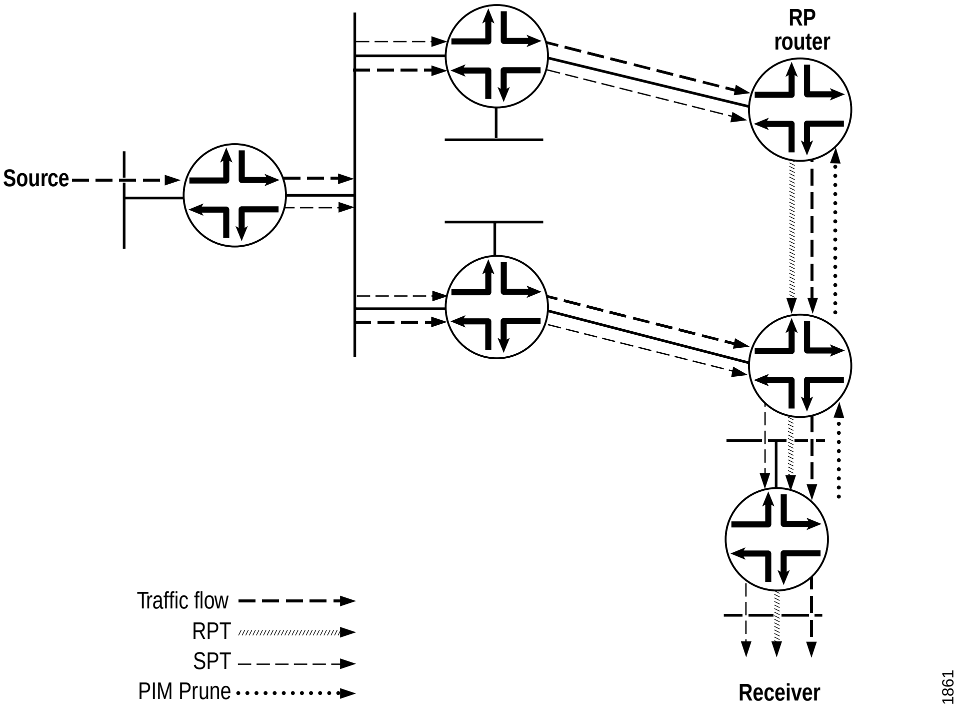
3. Multicast packets from that particular source begin coming from the source's DR and flowing down the new SPT to the receiver's DR. The receiver's DR is now receiving two copies of each multicast packet sent by the source—one from the RPT and one from the new SPT.

Figure 16: Receiver DR Sends a PIM Join to the Source



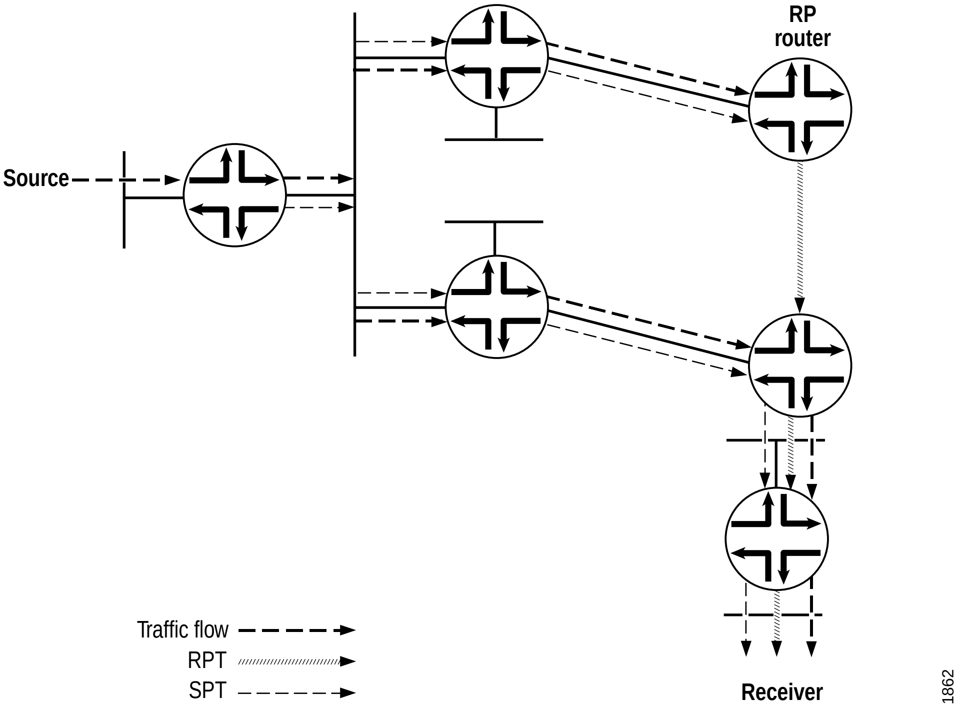
4. To stop duplicate multicast packets, the receiver’s DR sends a PIM Prune message towards the RP router, letting it know that the multicast packets from this particular source coming in from the RPT are no longer needed (see Figure 17).

Figure 17: PIM Prune Message is Sent from the Receiver’s DR Towards RP Router



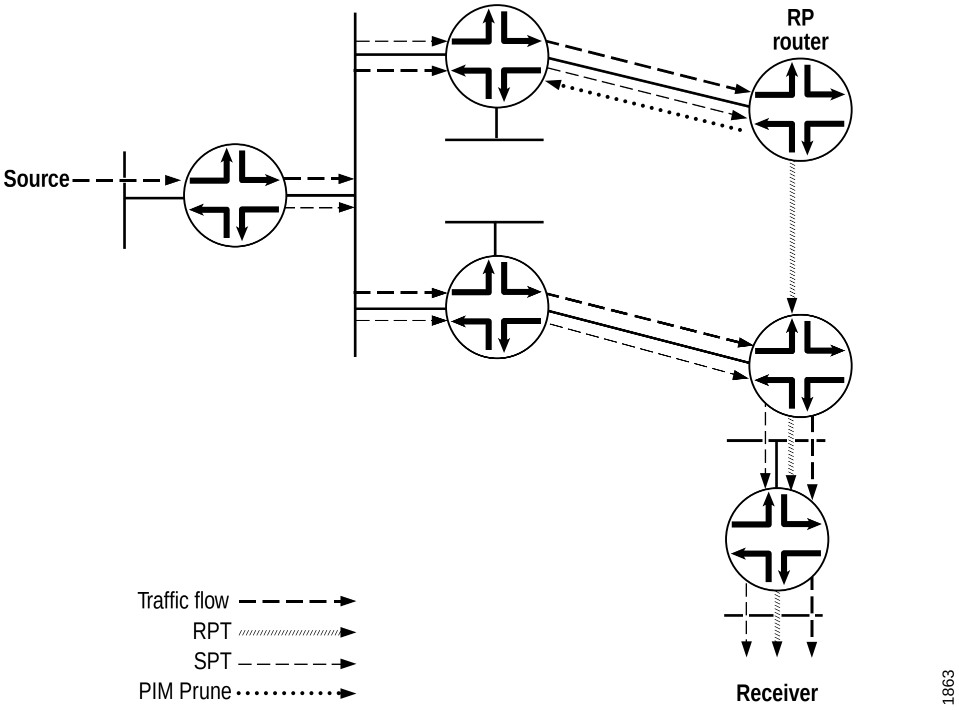
5. The PIM Prune message is received by the RP router, and it stops sending multicast packets down to the receiver’s DR. The receiver’s DR is only getting multicast packets for this particular source over the new SPT. However, multicast packets from the source are still arriving from the source’s DR towards the RP router (see Figure 18).

Figure 18: RP Router Receives PIM Prune Message



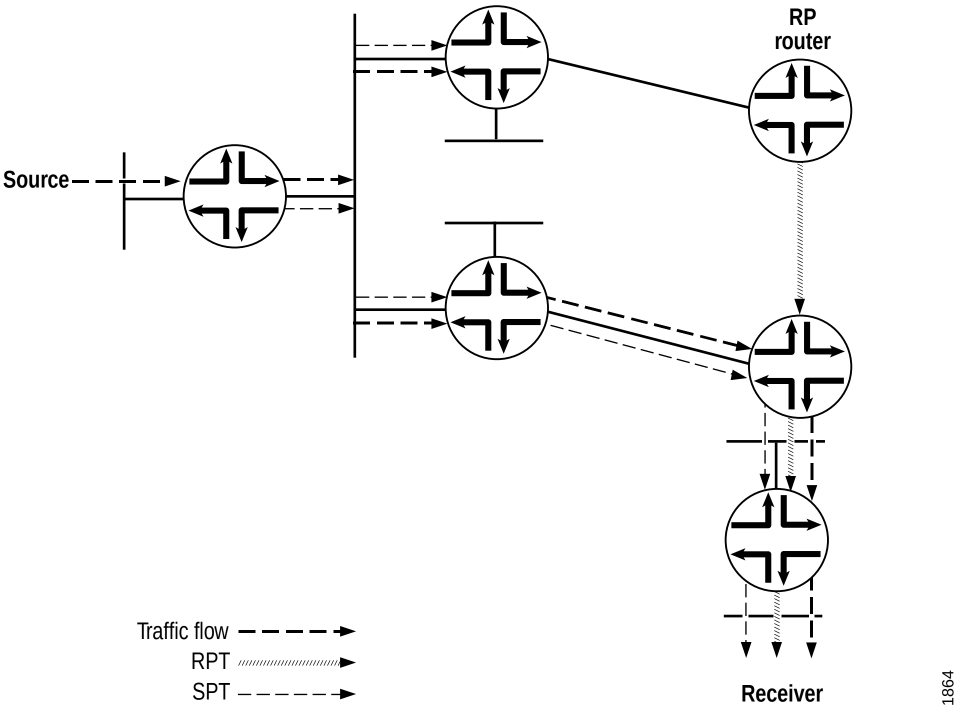
6. To stop the unneeded multicast packets from this particular source, the RP router sends a PIM Prune message towards the source’s DR (see Figure 19).

Figure 19: RP Router Sends a PIM Prune Message to Source’s DR



7. The receiver's DR now only receives multicast packets for the particular source from the SPT (see Figure 20).

Figure 20: Source's DR Stops Sending Duplicate Multicast Packets Toward RP Router



PIM-SSM

PIM-SSM is an extension of the PIM protocol. Using SSM, a client can receive multicast traffic directly from the source. PIM-SSM uses the PIM-SM functionality to create an SPT between the receiver and the source, but builds the SPT without the help of an RP.

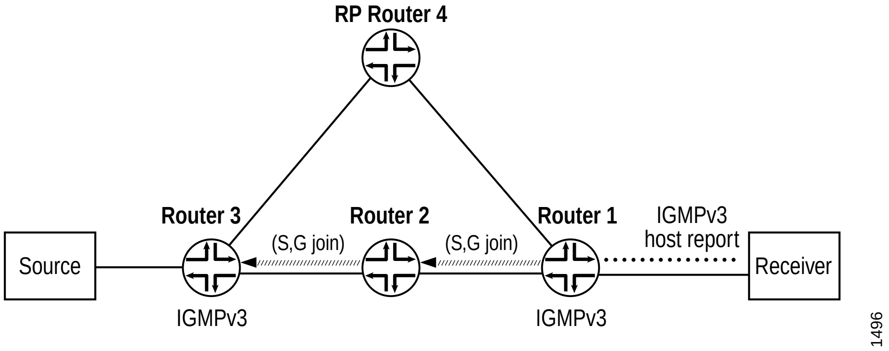
By default, the SSM group multicast address is limited to the IP address range 232.0.0.0 to 232.255.255.255. However, you can extend SSM operations into another Class D range by including the address statement at the [edit routing-options multicast ssm-groups] hierarchy level.

An SSM-configured network has distinct advantages over a traditionally configured PIM-SM network. There is no need for shared trees or RP mapping (no RP is required), or for RP-to-RP source discovery through Multicast Source Discovery Protocol (MSDP).

Deploying SSM is easy. You need only configure PIM-SM on all router interfaces and issue the necessary SSM commands, including specifying IGMP version 3 on the receiver's LAN. If PIM-SM is not explicitly configured on both the source and group members interfaces, multicast packets will not be forwarded. Source lists, supported in IGMP version 3, are used in PIM-SSM. As sources become active and start sending multicast packets, interested receivers in the SSM group receive the multicast packets.

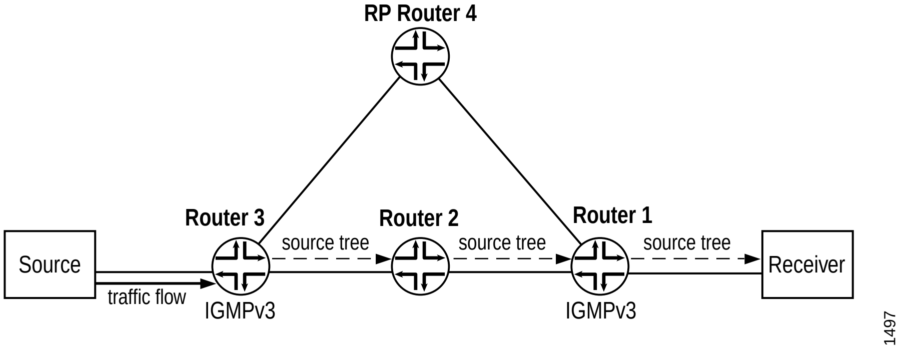
In a PIM-SSM-configured network, a host subscribes to an SSM channel (by means of IGMP version 3), announcing a desire to join group G and source S (see Figure 21). The directly connected PIM-SM router, the receiver's designated router (DR), sends an (S,G) Join message to its RPF neighbor for the source. Notice in Figure 21 that the RP is not contacted in this process by the receiver, as would be the case in normal PIM-SM operations.

Figure 21: Receiver Announces Desire to Join Group G and Source S



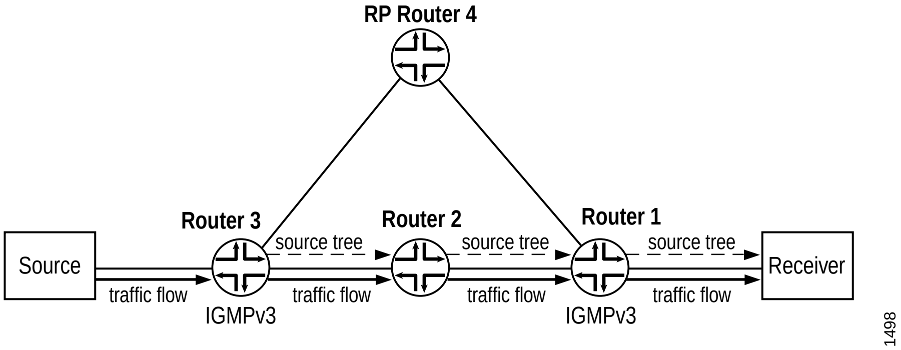
The (S,G) Join message initiates the source tree, then builds it out hop by hop until it reaches the source. In Figure 22, the source tree is built across the network to Router 3, the last-hop router connected to the source.

Figure 22: Router 3 (Last-hop Router) Joins the Source Tree



Using the source tree, multicast traffic is delivered to the subscribing host (see Figure 23).

Figure 23: The (S,G) State is Built Between the Source and the Receiver



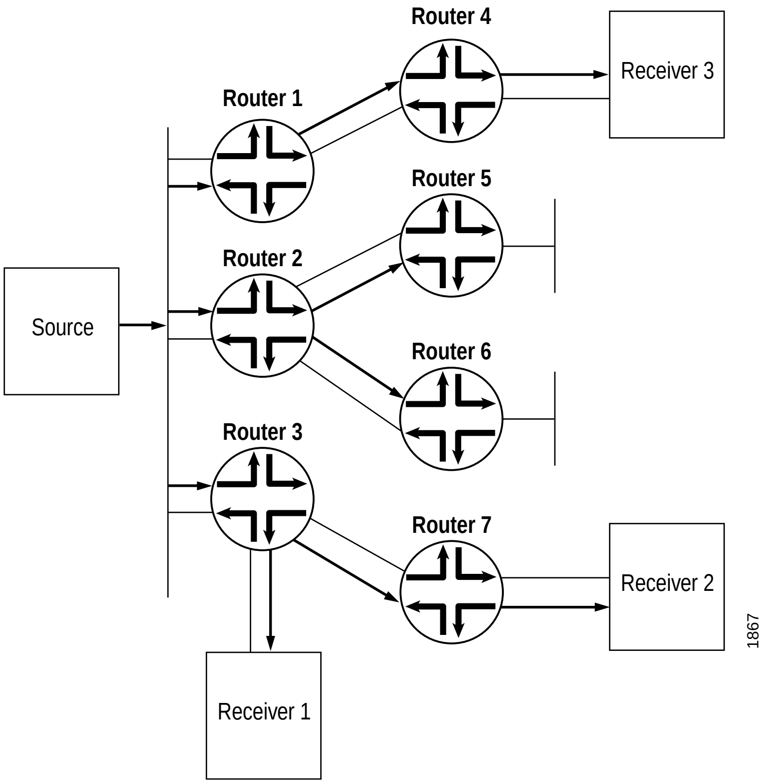
To configure additional SSM groups, include the ssm-groups statement at the [edit routing-options multicast] hierarchy level.

For more information about PIM-SSM, see “Example: Configure PIM-SSM on a Network” on page 65.

PIM Dense Mode

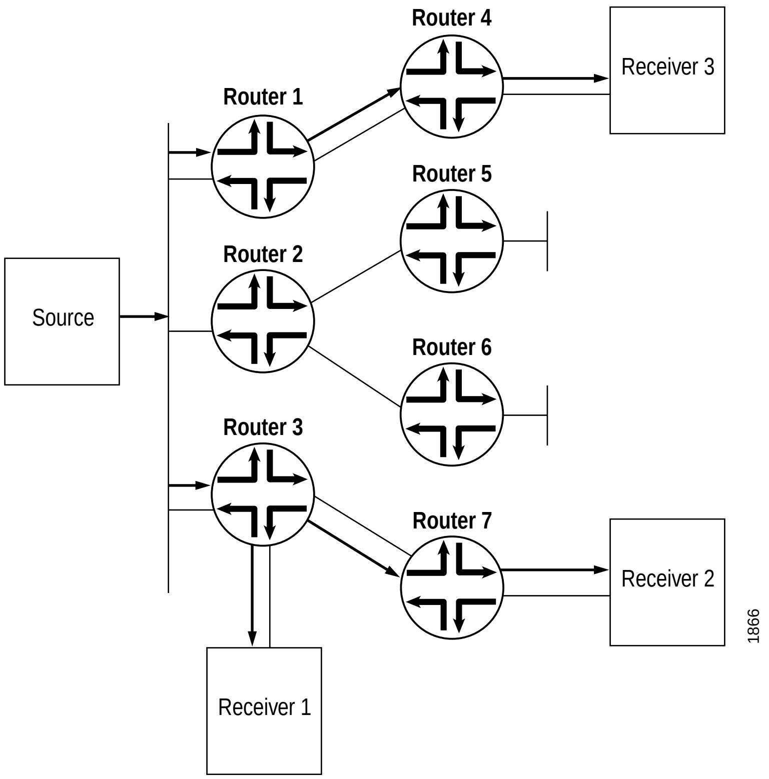
Unlike sparse mode, in which data is forwarded only to routers sending an explicit request, dense mode implements a *flood-and prune* mechanism, similar to the Distance Vector Multicast Routing Protocol (DVMRP). In PIM dense mode (PIM-DM), there is no RP. A router receives the multicast data on the interface (IIF) closest to the source, then forwards the traffic to all other interfaces. (see Figure 24).

Figure 24: Multicast Traffic Flooded From the Source Using PIM-DM



Flooding occurs periodically. It is used to refresh state information, such as the source IP address and multicast group pair. If the router has no interested receivers for the data, and the OIL becomes empty, the router sends a Prune message upstream to stop delivery of multicast traffic (see Figure 25).

Figure 25: Prune Messages Sent Back to the Source to Stop Unwanted Multicast Traffic



PIM Sparse-Dense Mode

Sparse-dense mode, as the name implies, allows the interface to operate on a per-group basis in either sparse or dense mode. A group specified as dense is not mapped to an RP. Instead, data packets destined for that group are forwarded by means of PIM-DM rules. A group specified as sparse is mapped to an RP, and data packets are forwarded by means of PIM-SM rules.

For information about PIM-SM and PIM-DM rules, see “PIM Sparse Mode” on page 93 and “PIM Dense Mode” on page 108.

RP Mapping Using Anycast RP

For the purposes of load balancing and redundancy, you can configure anycast RP. When you configure anycast RP, you bypass the restriction of having one active RP per multicast group, and instead deploy multiple RPs for the same group range. The RP routers share one unicast IP address. Sources from one RP are known to other RPs that use MSDP. Sources and receivers use the closest RP, as determined by the IGP.

You can use anycast RP within a domain to provide redundancy and RP load sharing. When an RP goes down, sources and receivers are taken to a new RP by means of unicast routing.

Anycast RP is defined in Internet draft draft-ietf-mboned-anycast-rp-08.txt, *Anycast RP Mechanism Using PIM and MSDP*. To access Internet RFCs and drafts, go to the IETF Web site at http://www.ietf.org.

We recommend a static RP mapping with anycast RP over a bootstrap router (BSR) and auto-RP configuration because it provides all the benefits of BSR and auto-RP without the complexity of the BSR and auto-RP mechanisms.

See also "Example: Configure Anycast RP" on page 133.

Multicast Over Layer 3 VPNs

In the unicast environment for Layer 3 virtual private networks (VPNs), all VPN state information is contained within the Provider Edge (PE) routers. However, with multicast for Layer 3 VPNs, PIM adjacencies are established in one of the following ways:

You can set PIM adjacencies between the Customer Edge (CE) router and the PE router through a virtual routing and forwarding (VRF) instance at the [edit routing-instances *instance-name* protocols pim] hierarchy level. You must include the new vpn-group-address statement at this hierarchy level, specifying a multicast group. The RP listed within the VRF-instance is the VPN-RP.

You can also set the Master PIM instance and the PE's IGP neighbors by configuring statements at the [edit protocols pim] hierarchy level. You must add the multicast group specified in the VRF instance to the master PIM instance. The set of master PIM adjacencies throughout the SP network make up the forwarding path that becomes an RP tree rooted at the SP-RP. Therefore, Provider (P) routers within the provider core must maintain multicast state information for the VPNs.

For this to work properly, you need two types of RP routers for each VPN:

A VPN-RP—an RP router located somewhere within the VPN

An SP-RP—an RP router located within the SP network

Note

A PE router can act as an SP-RP, but the PE router cannot be the VPN-RP of a VPN. The VPN-RP must be located at the customer edge, or somewhere else within the VPN.

For more information about configuring multicast for Layer 3 VPNs, see "Configure Multicast for Layer 3 VPNs" on page 126. For multicast Layer 3 VPN examples, see "Example: Configure PIM-SM Multicast Over Layer 3 VPNs" on page 137.

Tunnel PICs and Multicast

On Juniper Networks routers, data packets are encapsulated and de-encapsulated into tunnels by means of hardware—not software running on the router's processor. The hardware used to create tunnel interfaces is a Tunnel Services PIC. All RP routers and PIM-SM DRs connected to a source require a Tunnel Services PIC.

In PIM-SM, the source DR takes the initial multicast packets and encapsulates them in PIM Register messages. It then unicasts them to the PIM-SM RP router, where the PIM Register message is de-encapsulated.

When a router is configured as a PIM-SM RP router (by specifying an address using the address statement at the [edit protocols pim rp local] hierarchy level) and a Tunnel PIC is present on the router, a PIM Register de-encapsulation interface, or pd interface, is automatically created. The pd interface receives PIM register messages and de-encapsulates them by means of the hardware.

If PIM-SM is enabled on any router (potentially a PIM-SM source DR) and a Tunnel Services PIC is present on the router, a PIM Register encapsulation interface, or pe interface, is automatically created for each RP address that is used to encapsulate source data packets and send them to respective RP addresses on PIM-DR as well as PIM-RP. The pe interface receives PIM register messages and encapsulates them by means of the hardware.

If the source DR *is* the RP, then there is no need for PIM Register messages and consequently no need for a Tunnel Services PIC to be present.

Fast Ethernet PIC

The Fast Ethernet PICs allow you to use a Juniper Networks router without having to purchase a Gigabit Ethernet switch. For example, with a Fast Ethernet PIC, local devices (such as a DNS server) could connect directly to a Juniper Networks router without a Gigabit Ethernet switch. You can use a Fast Ethernet PIC for private peering between two colocated routers.

When a router is truly a "multicast router" on an interface/port location, it sets the interface MAC filter to multicast promiscuous mode, and the number of multicast groups are unlimited. However, when the router is not exclusively used for multicasting and other protocols such as OSPF, RIPv2, or NTP are running, they individually ask the interface to program the MAC filter to pick up their respective multicast group alone and the maximum number of multicast MAC filters is limited to 20. In example, the maximum number of multicast MAC filters for protocols such as OSPF with group 224.0.0.5 is 20.

Filtering

Along with applying MSDP SA filters on all external MSDP sessions (in and out) to prevent SAs for groups and sources from leaking in and out of the network, you need to apply BSR filters. Applying a BSR filter to the boundary of a network prevents foreign BSR messages (which announce RP addresses) from leaking into your network. Since the routers in a PIM-SM domain should only know the address of one RP router, having more than one in the network can create problems. See “Example: Configure PIM BSR Filters” on page 135 for a sample filter configuration.

If you did not use multicast scoping to create boundary filters for all customer-facing interfaces, you might want to use PIM Join filters. Multicast scopes prevent the actual multicast data packets from flowing in or out of an interface. PIM join filters prevent PIM-SM state from being created in the first place. Since PIM join filters only apply to PIM-SM state, it might be more beneficial to use multicast scoping to filter the actual data.

For more information, see “Multicast Scoping Overview” on page 57 and “Example: Configure PIM Join Filters” on page 135.

Note

When applying firewall filters, firewall action modifiers, such as log, sample, and count, work only when you apply the filter on an inbound interface. The modifiers will not work on an outbound interface.